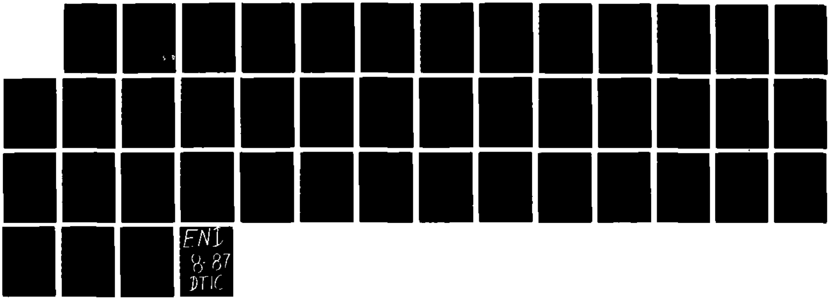
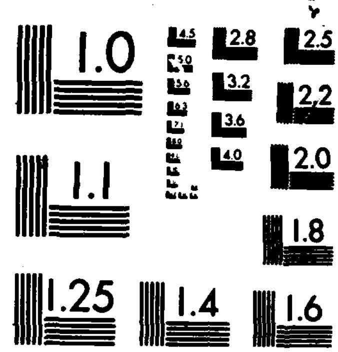
AD-A182 552 ARMA ESTIMATORS OF PROBABILITY DENSITIES WITH EXPONENTIAL OR REGULARLY VA..(U) TEXAS A AND M UNIV COLLEGE STATION DEPT OF STATISTICS J D HART JUN 87 1/1

UNCLASSIFIED TR-2 N00014-85-K-0723 F/G 12/3 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

Department of
STATISTICS
Phone 409 - 845-3141

ARMA ESTIMATORS OF PROBABILITY DENSITIES WITH EXPONENTIAL OR REGULARLY VARYING FOURIER COEFFICIENTS

Jeffrey D. Hart
Department of Statistics
Texas A&M University

Technical Report No. 2

June 1987

Texas A&M Research Foundation
Project No. 5321

'Nonparametric Estimation of Functions Based Upon Correlated Observations'

Sponsored by the Office of Naval Research

Dr. Jeffrey D. Hart and Dr. Thomas E. Wehrly
Co-Principal Investigators

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2
2. GOVT ACCESSION NO.	ADA182552
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle)	ARMA Estimators of Probability Densities with Exponential or Regularly Varying Fourier Coefficients
5. TYPE OF REPORT & PERIOD COVERED	Technical
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s)	Jeffrey D. Hart
8. CONTRACT OR GRANT NUMBER(s)	N00014-85-K-0723
9. PERFORMING ORGANIZATION NAME AND ADDRESS	Texas A&M University Department of Statistics College Station, TX 77843
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS	
11. CONTROLLING OFFICE NAME AND ADDRESS	Office of Naval Research 800 North Quincy Street Arlington, Virginia 2217-50000
12. REPORT DATE	June 1987
13. NUMBER OF PAGES	43
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	
15. SECURITY CLASS. (of this report)	Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Probability density estimation, Fourier series, generalized jackknife, regularly varying function.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Properties of a probability density estimator having the rational form of an ARMA spectrum are investigated. Under various conditions on the underlying density's Fourier coefficients, the ARMA estimator is shown to have asymptotically smaller mean integrated squared error (MISE) than the best window-type Fourier series estimator. The most interesting cases are those in which the Fourier coefficients are regularly varying with index $-\rho, \rho>1/2$. For example, when $\rho=2$ the asymptotic MISE of a certain ARMA estimator is only about 75% of that for the optimum window estimator. For a density f with support in $[0,\pi]$, the condition $\rho=2$ occurs whenever $f'(0+)\neq 0$, $f'(\pi-)=0$, and f'' is square integrable.

1

ARMA ESTIMATORS OF PROBABILITY DENSITIES WITH EXPONENTIAL OR REGULARLY VARYING FOURIER COEFFICIENTS

Jeffrey D. Hart

Department of Statistics
Texas A&M University

Properties of a probability density estimator having the rational form of an ARMA spectrum are investigated. Under various conditions on the underlying density's Fourier coefficients, the ARMA estimator is shown to have asymptotically smaller mean integrated squared error (MISE) than the best window-type Fourier series estimator. The most interesting cases are those in which the Fourier coefficients are regularly varying with index $-\rho$, $\rho > 1/2$. For example, when $\rho = 2$ the asymptotic MISE of a certain ARMA estimator is only about 75% of that for the optimum window estimator. For a density f with support in $[0, \pi]$, the condition $\rho = 2$ occurs whenever $f'(0+) \neq 0$, $f'(\pi-) = 0$, and f'' is square integrable.

Key words and phrases. Probability density estimation, Fourier series, generalized jackknife; regularly varying function.

1. Introduction

Suppose $X_1,\ldots,X_n$ are independent observations from a density f with support contained in $[0,\pi]$. We consider estimating f(x) by a quantity having the rational form

$$\hat{f}(x) = (\hat{\beta}_0 + 2 \sum_{j=1}^{k} \hat{\beta}_j \cos jx)/|1-\alpha e^{ix}|^2, \quad (1.1)$$

where $|\alpha|<1$ and the $\hat{\beta}_j$'s are explicit functions of $X_1,\ldots,X_n$. An estimator as in (1.1) bears an obvious resemblance to autoregressive moving average (or ARMA) spectra, and will thus be referred to as an ARMA estimator.

Apparently, ARMA type probability density estimators have not previously been considered in the statistical literature. Parzen (1979) and Carmichael (1984) have, however, proposed autoregressive (or AR) type estimators of the form

$$\hat{f}(x;p) = \hat{c}|1-\hat{\alpha}_1 e^{ix}-\ldots-\hat{\alpha}_p e^{ipx}|^{-2}.$$

Carmichael (1984) obtains a consistency result for $\hat{f}(\cdot;p)$ by allowing the AR order, p, to tend to infinity at a certain rate with the sample size, n. In the current paper, the AR order is fixed at 1 and the MA order tends to infinity with n. The motivation for the latter scheme is based on the numerical analytic device known as the e_1-transform, which in turn is related to the notion of a generalized jackknife.

Hart and Gray (1985) studied the use of ARMA representations in approximating (rather than estimating) density functions. Their results characterize the integrated squared bias of (1.1) in a number of different situations. The current paper greatly generalizes the results of Hart and Gray and also considers the mean integrated squared error (MISE) of (1.1). It is shown that there exist quite general conditions under which an estimator of the form (1.1) has asymptotically smaller MISE than does <u>any</u> window-type Fourier series estimator. For example, this optimality property obtains when $f'(0+)\neq 0$, $f'(\pi-)=0$, and f'' is square integrable. More generally, we obtain results on the behavior of (1.1)'s MISE when the Fourier coefficients of f are either approximately exponential or regularly varying at infinity.

The paper will be ordered in the following way. In Section 2, the estimator to be studied is defined, and a number of motivations for its use are given. Section 3 contains some basic results concerning the MISE of the ARMA estimator. The asymptotic MISE of the estimator is studied in Sections 4 and 5 under the aforementioned conditions on the Fourier coefficients of f. In Section 6 it is shown that when f is smooth (in a well-defined sense) at one endpoint of its support but not at the other, then the ARMA estimator is asymptotically superior to traditional window-type estimators. Cross-validated smoothing of ARMA estimates is addressed in Section 7, and some summary remarks are given in Section 8.

2. The Proposed Estimator

Let $X_1,\ldots,X_n$ be a random sample from a density f with support contained in $[0,\pi]$. We shall assume that f has the Fourier series

$$f(x) \sim \pi^{-1}\left(1+2\sum_{j=1}^{\infty}\phi_j\cos jx\right),\ 0 \le x \le \pi, \tag{2.1}$$

where $\phi_j = \int_0^\pi \cos jx\, f(x)dx$. Cencov (1962), Kronmal and Tarter (1968), Hall (1983) and others have investigated density estimators of the form

$$\hat{f}_n(x;m) = \pi^{-1}\left(1+2\sum_{j=1}^{m}\hat{\phi}_j\cos jx\right),\ 0 \le x \le \pi,$$

where

$$\hat{\phi}_j = n^{-1}\sum_{k=1}^{n}\cos jX_k.$$

The estimators to be studied here are

$$f_n(x;m,\alpha) = f_n(x;m) + (2/\pi)\,\mathrm{Real}\left[\frac{\hat{\phi}_m\alpha\exp(i(m+1)x)}{1-\alpha\exp(ix)}\right], \tag{2.2}$$

where $-1 < \alpha < 1$. The pair (m,α) is the smoothing parameter of the estimator and can be chosen from the data by cross-validation. We will return to this point in Section 7.

Before further discussion of the proposed estimator, we should justify using the cosine basis as opposed to a basis with both cosine and sine functions. This study was in part motivated by the problem of estimating animal abundance using the line transect method (see Gates and Smith 1980 and Crain, Burnham, Anderson, and Laake 1979). The density function estimated in this setting is typically assumed to be monotone decreasing on $(0,\pi)$. Since $f(0)\neq f(\pi)$, the periodic extension of f is discontinuous at 0 and π, and hence a cosine-sine Fourier series estimator will perform poorly near these two points. As shown by Hall (1983), though, the cosine series estimator is not adversely affected by the condition $f(0) \neq f(\pi)$. In kernel estimation, the analogous means of correcting boundary problems is the symmetrization device studied by Schuster (1985).

There are a number of ways of characterizing the estimators $\hat{f}_n(\cdot;m,\alpha)$. First, it is clear that they may be written

$$\hat{f}_n(x;m,\alpha) = \sum_{j=-(m+1)}^{m+1} \hat{\beta}_j e^{ijx} / |1-\alpha e^{ix}|^2, 0 \leq x \leq \pi, \tag{2.3}$$

where $\hat{\beta}_j = \hat{\beta}_{-j}$ and the $\hat{\beta}_j$'s depend only on α and the $\hat{\phi}_j$'s. Aside from the issue of positivity, (2.3) has the form of an ARMA(1,m+1) spectrum, hence the name ARMA probability density estimator. The form (2.3) suggests that $\hat{f}_n(\cdot;m,\alpha)$ is well suited for densities with large "power" at either 0 or π (but not both). While this is so, it will be seen in Section 6 that such an observation somewhat understates the value of ARMA estimators.

Perhaps a more interesting way of characterizing $\hat{f}_n(\cdot,m,\alpha)$ is in terms of the generalized jackknife and the numerical analytic device known as the e_1-transform. Using (2.2), it is easy to show that, for $m \geq 1$,

$$\hat{f}_n(x;m,\alpha) = \pi^{-1}\left[1 + 2\,\mathrm{Real}\left(\frac{\hat{F}_m(x) - \alpha e^{ix}\,\hat{F}_{m-1}(x)}{1 - \alpha e^{ix}}\right)\right]. \tag{2.4}$$

where $\hat{F}_k(x) = \sum_{j=1}^{k} \hat{\phi}_j e^{ijx}$, k=1,2,..., and $\hat{F}_0 \equiv 0$. The quantity $G_{m,\alpha}(x) = (\hat{F}_m(x) - \alpha e^{ix} \hat{F}_{m-1}(x))/(1 - \alpha e^{ix})$ is a generalized jackknife estimator (as defined by Schucany, Gray, and Owen 1971) of the function

$$F(x) = \sum_{j=1}^{\infty} \phi_j e^{ijx} .$$

The problem of choosing α in such a way that $G_{m,\alpha}(x)$ has smaller bias than $\hat{F}_m(x)$ will be addressed later.

Defining $F_k(x) = E(\hat{F}_k(x))$, we have

$$E(G_{m,\alpha}(x)) = (F_m(x) - \alpha e^{ix} F_{m-1}(x))/(1 - \alpha e^{ix}).$$

If α is taken to be ϕ_{m+1}/ϕ_m, then $\{E(G_{m,\alpha}(x)):m=1,2,...\}$ is the e_1-transform of the sequence $\{F_m(x)\}$. This transform, which dates at least to Aitken (1926), is a numerical analytic tool used for accelerating the convergence of a sequence to its limit. The work of Shanks (1955) gives a thorough account of e_1 and the more general e_n-transform. The interesting and enlightening paper of H. L. Gray (1985) demonstrates the close connection between many numerical analytic methods (including the e_n-transform) and the statistical notion of jackknifing to reduce bias. For more uses of the e_n-transform in statistical problems, see Gray, Watkins, and Adams (1972) and Morton and Gray (1984).

Also of interest is the Fourier series of $\hat{f}_n(\cdot;m,\alpha)$. From (2.2),

$$\hat{f}_n(x;m,\alpha) = \hat{f}_n(x;m) + (2/\pi) \sum_{j=m+1}^{\infty} \hat{\phi}_m \alpha^{j-m} \cos jx .$$

The Fourier coefficients, $\hat{\phi}_j(m,\alpha)$, of the ARMA estimator are thus

$$\hat{\phi}_j(m,\alpha) = \begin{cases} \hat{\phi}_j & , j = 0,1,\ldots,m \\ \hat{\phi}_m \alpha^{j-m}, & j = m+1,\ldots \end{cases} .$$

This shows that $\hat{f}_n(x;m,\alpha)$ will tend to have smaller bias than $\hat{f}_n(x;m)$ when the ϕ_j decay geometrically. As will be seen, though, geometrically decaying ϕ_j's are but a subset of the cases in which $\hat{f}_n(\cdot;m,\alpha)$ affords a bias reduction. Hart and Gray (1985) argue that ARMA type approximators are very generally an effective means of counteracting the leakage effect inherent in $\hat{f}_n(\cdot;m)$.

3. Mean Integrated Squared Error of $\hat{f}_n(\cdot;m,\alpha)$

Define the mean integrated squared error of the estimator $\hat{f}$ by $J(\hat{f},f) = E\int_0^\pi (\hat{f}(x)-f(x))^2dx$. The MISE will be used as a basis for comparing $\hat{f}_n(\cdot;m,\alpha)$ with $\hat{f}_n(\cdot;m)$ and more general Fourier series estimators.

It is straightforward to show that

$$J(\hat{f}_n(\cdot;m,\alpha),f) = (2/\pi)\,[n^{-1}\sum_{j=1}^{m}\mathrm{var}(\cos jX_1) + \alpha^2(1-\alpha^2)^{-1}\,\mathrm{var}(\cos mX_1)/n]$$

$$+ (2/\pi)\sum_{j=m+1}^{\infty}(\phi_j-\phi_m\alpha^{j-m})^2 \quad , \tag{3.1}$$

where $\mathrm{var}(\cos jX_1) = (1+\phi_{2j})/2 - \phi_j^2$. The first of the two terms in (3.1) is the integrated variance of $\hat{f}_n(\cdot;m,\alpha)$ while the latter is the integrated squared bias. Note that taking $\alpha=0$ in (3.1) yields $J(\hat{f}_n(\cdot;m),f)$. This is important since it makes clear that one may always choose α so that the MISE of an ARMA estimator is no bigger than that of $\hat{f}_n(\cdot;m)$.

If $\{\alpha_{m,n}:m,n = 1,2,\ldots\}$ is a sequence satisfying $|\alpha_{m,n}|<1$ and $\liminf_{n,m\to\infty} (1-\alpha^2_{m,n})>0$, then as $n,m\to\infty$

$$J(\hat{f}_n(\cdot;m,\alpha_{m,n}),f) = (2/\pi)[n^{-1}\sum_{j=1}^{m} \operatorname{var}(\cos jX_1) + \sum_{j=m+1}^{\infty}(\phi_j-\phi_m\alpha_{m,n}^{j-m})^2] + O(n^{-1}). \tag{3.2}$$

In this event, then, the effect of $\alpha_{m,n}$ on the variance of the ARMA estimator is asymptotically negligible as n and $m\to\infty$. It will be seen, though, that there are also cases where a good choice of $\alpha_{m,n}$ is such that $m(1-\alpha^2_{m,n})\to c$ as $m,n\to\infty$. In such cases the term $\alpha^2_{m,n}(1-\alpha^2_{m,n})^{-1}\operatorname{var}(\cos mX_1)/n$ is of order m/n and cannot be ignored in asymptotic considerations.

Before proceeding, we give the following useful lemma.

<u>Lemma 1</u>. Suppose f has square summable Fourier coefficients. Then, as $m\to\infty$,

$$\sum_{j=1}^{m} \operatorname{var}(\cos jX_1) = m/2 + O(\sqrt{m}).$$

If it is further required that the ϕ_j's be absolutely summable the term $O(\sqrt{m})$ may be replaced by $O(1)$.

Proof: Recall that

$$\sum_{j=1}^{m} \operatorname{var}(\cos jX_1) = m/2 + (1/2)\sum_{j=1}^{m} \phi_{2j} - \sum_{j=1}^{m} \phi_j^2.$$

If the ϕ_j's are absolutely summable, it is clear that the last expression is $m/2 + O(1)$. By the Cauchy-Schwarz inequality,

$$\left|\sum_{j=1}^{m} \phi_{2j}\right| \le \sqrt{m}\left(\sum_{j=1}^{m} \phi_{2j}^2\right)^{1/2},$$

from which the rest of the lemma follows.

4. Densities With Nearly Exponential Fourier Coefficients

In Section 2 it was mentioned that the ARMA estimator should perform well for densities with geometrically decaying ϕ_j's, i.e., ones for which $\phi_j \sim c\exp(-aj)$ as $j\to\infty$. In this section, we shall investigate the more general case in which f has Fourier coefficients

$$\phi_j = e^{-aj}R_\rho(j),\ j = 1,2,\ldots, \tag{4.1}$$

where $a>0$ and R_ρ is a regularly varying function (defined on $[1,\infty)$) with index $\rho, |\rho|<\infty$.

An ideal asymptotic comparison of $J(\hat{f}_n(\cdot;m,\alpha),f)$ and $J(\hat{f}_n(\cdot;m),f)$ would be to investigate $R_n = J_n^*/J_n$, where $J_n^* = \min_{(m,\alpha)} J(\hat{f}_n(\cdot;m,\alpha),f)$ and $J_n = \min_m J(\hat{f}_n(\cdot;m),f)$. Instead of this approach, we shall choose α to depend on at most m, and then study $\min_m J(\hat{f}_n(\cdot;m,\alpha_m),f)$ as $n\to\infty$. The α_m to be used is one suggested by generalized jackknife theory. According to Schucany, Gray, and Owen (1972),

$$E\left((\hat{F}_m(x) - r\hat{F}_{m-1}(x))/(1-r)\right) = \sum_{j=1}^{\infty} \phi_j e^{ijx}$$

if $r = \sum_{j=m+1}^{\infty} \phi_j e^{ijx} / \sum_{j=m}^{\infty} \phi_j e^{ijx}$. When (4.1) holds, the last quantity is asymptotic to $(\phi_{m+1}/\phi_m)e^{ix}$ as $m\to\infty$. Referring to (2.4), then, a reasonable choice for α_m would seem to be ϕ_{m+1}/ϕ_m. Recall that with $\alpha = \phi_{m+1}/\phi_m$, $E[G_{m,\alpha}(x)]$ is the e_1-transform of $F_m(x)$ (see Section 2).

Defining $B(\hat{f},f)$ to be the integrated squared bias of $\hat{f}$, it follows from Theorem 4 of Hart and Gray (1985) that whenever $\alpha_m = \phi_{m+1}/\phi_m \to e^{-a}(a>0)$

$$\lim_{m\to\infty} B(\hat{f}_n(\cdot;m,\alpha_m),f)/B(\hat{f}_n(\cdot;m),f) = 0.$$

When the ϕ_j satisfy (4.1) we have the following more precise result. As a matter of convenience, α_m is taken to be e^{-a} in the remainder of this section.

<u>Lemma 2.</u> Let f be a density with Fourier coefficients as in (4.1). Then, as $m\to\infty$,

$$B(\hat{f}_n(\cdot;m),f) \sim (2/\pi)e^{-2a(m+1)}R_\rho^2(m)(1-e^{-2a})^{-1},$$

$$B(\hat{f}_n(\cdot;m,e^{-a}),f)/B(\hat{f}_n(\cdot;m),f) \sim A_m =$$

$$(1-e^{-2a}) \sum_{j=0}^{\infty} (R_\rho(j+m+1)/R_\rho(m)-1)^2 e^{-2aj}, \text{ and}$$

$$\lim_{m\to\infty} A_m = 0.$$

Proof: Define

$$S_m = \sum_{j=m+1}^{\infty} \phi_j^2/(e^{-2a(m+1)}R_\rho^2(m))$$

$$= \sum_{j=0}^{\infty} e^{-2aj} R_\rho^2(j+m+1)/R_\rho^2(m).$$

Since R_ρ is regularly varying, it follows that $R_\rho^2(m+1)/R_\rho^2(m)\to 1$ as $m\to\infty$.

Hence, for $0<\varepsilon<e^{2a}-1$, all m sufficiently large and each $j\geq 0$

$e^{-2aj}R_\rho^2(j+m+1)/R_\rho^2(m) \leq e^{-2aj}(1+\varepsilon)^{j+1} = (1+\varepsilon)\exp[-j(2a-\log(1+\varepsilon))]$. Since the last expression is summable, dominated convergence immediately gives $S_m \to (1-e^{-2a})^{-1}$, proving the first part of the lemma. The rest of the lemma is easily obtained by taking $\alpha = e^{-a}$ in $B(\hat{f}_n(\cdot;m,\alpha), f)$ (see (3.1)) and arguing as we just did.

Under the conditions of Lemma 2, we see that the integrated squared bias of $\hat{f}_n(\cdot;m,e^{-a})$ is asymptotically negligible compared to that of $\hat{f}_n(\cdot;m)$. Furthermore, appealing to (3.2), it follows that $J(\hat{f}_n(\cdot;m,e^{-a}),f)$ is asymptotically no larger than $J(\hat{f}_n(\cdot;m),f)$ as $n,m\to\infty$. The following theorem gives a more precise result concerning the MISE's of the two estimators.

<u>Theorem 1.</u> Let R_ρ be any regularly varying function (defined on $[1,\infty)$) with the representation, for all t sufficiently large,

$$R_\rho(t) = t^\rho \exp(c + \int_B^t \frac{\varepsilon(x)}{x}\,dx),$$

where $\rho\neq 0$, $|c|<\infty$, and $B>1$. If the Fourier coefficients of f are

$$\phi_j = e^{-aj}R_\rho(j) \quad , j = 1,2,\ldots,$$

then

$$J(\hat{f}_n(\cdot;m),f) = (1/\pi)\,[m/n+2\phi_m^2(e^{2a}-1)^{-1}] + O(1/n) + o(\phi_m^2) \qquad (4.2)$$

and

$$J(\hat{f}_n(\cdot;m,e^{-a}),f) = (1/\pi)\,[m/n+2\rho^2\phi_m^2\, m^{-2}C_a]$$

$$+ O(1/n) + o(\phi_m^2\, m^{-2}), \qquad (4.3)$$

where $C_a = e^{-2a}(1+e^{-2a})(1-e^{-2a})^{-3}$. Defining m_n to be the minimizer of $J(\hat{f}_n(\cdot;m),f)$ and m_n^* to be the minimizer of $m/n + 2\rho^2\phi_m^2 m^{-2}C_a$, we also have

$$J(\hat{f}_n(\cdot;m_n),f) = (2a\pi n)^{-1}\,[\log n + 2\rho\log\log n] + o(\log\log n/n),$$

and

$$J(\hat{f}_n(\cdot;m_n^*,e^{-a}),f) = (2a\pi n)^{-1}\,[\log n + 2(\rho-1)\log\log n]$$
$$+\, o(\log\log n/n).$$

Proof: Equation (4.2) is an immediate consequence of Lemmas 1 and 2. To obtain (4.3), note that for all m sufficiently large

$$R_\rho(j+m+1)/R_\rho(m)-1 =$$

$$(1+(j+1)/m)^\rho \exp\left(\int_m^{m+j+1} \varepsilon(x)x^{-1}\,dx\right) - 1 =$$

$$(1+(j+1)/m)^\rho - 1 + (1+(j+1)/m)^\rho\left[\exp\left(\int_m^{m+j+1} \varepsilon(x)x^{-1}\,dx\right) - 1\right]$$

$$= \rho b_{j,m}^{\rho-1}(j+1)m^{-1} + (1+(j+1)/m)^\rho\left[\exp(c_{j,m})\int_m^{m+j+1} \varepsilon(x)x^{-1}\,dx\right],$$

where $1 \le b_{j,m} \le 1 + (j+1)/m$ and $c_{j,m}$ is a number between 0 and $\int_m^{m+j+1} \varepsilon(x)x^{-1}\,dx$.

We have

$$\left|\int_m^{m+j+1} \varepsilon(x)x^{-1}\,dx\right| \le m^{-1}(j+1)\sup_{m\le x\le m+j+1}|\varepsilon(x)| \quad ,$$

and since R_ρ is regularly varying, $\varepsilon(x)\to 0$ as $x\to\infty$ (see Seneta 1976, p. 3). Applying Lemma 2, the above results, and dominated convergence gives

$$B(\hat{f}_n(\cdot;m,e^{-a}),f) \sim (2/\pi)e^{-2a}\phi_m^2\, m^{-2}\rho^2 \sum_{j=0}^{\infty} (j+1)^2 e^{-2aj}.$$

Equation (4.3) follows from this and Lemma 1. For the rest of the proof, note that m_n is such that $\Delta(n,m_n)\le 0$ and $\Delta(n,m_n+1)\ge 0$, where

$$\Delta(n,t) = (1+R_\rho(2t)e^{-2at}) - 2(n+1)R_\rho^2(t)e^{-2at}.$$

Since $\Delta(n;\cdot)$ is eventually continuous, it follows that for all n sufficiently large $\Delta(n,t_n) = 0$ for some $t_n \,\varepsilon[m_n,m_n+1]$. Now $\Delta(n,t_n) = 0$ implies

$$t_n[1-\rho\log t_n/(at_n) - (c+\int_B^{t_n} \varepsilon(x)x^{-1}\,dx)/(at_n)]$$

$$=(2a)^{-1}\log(n+1)[1-\log(1/2 + R_\rho(2t_n)\exp(-2at_n)/2)/\log (n+1)].$$

This implies that $t_n = (2a)^{-1} \log n + o(\log n)$. Substitution of this expression for t_n into the previous equation gives

$$t_n = (2a)^{-1}[\log n + 2\rho\log\log n + o(\log\log n)],$$

where we make use of the fact that $\log L(x)/\log x\to 0$ for any slowly varying function L (see Seneta, p. 18). The expansion for $J(\hat{f}_n(\cdot;m_n),f)$ follows upon observing that $m_n = t_n+O(1)$ and using (4.2). The proof for the ARMA estimator follows in an analogous manner.

Theorem 1 gives conditions under which an ARMA estimator has, for all n bigger than some n_0, smaller MISE than does the best estimator $\hat{f}_n(\cdot,m)$. The improvement, though, is only in terms of second order efficiency. This occurs for two reasons. First, for the ϕ_j's considered, the two optimum MISEs are dominated by their variance terms, and so the smaller bias of the ARMA estimator is not reflected in the leading MISE term. Secondly, ϕ_j's of the form (4.1) with $\rho \neq 0$ are not near enough to being geometric for the ARMA estimator to be fully effective. The following theorem provides conditions under which an ARMA estimator yields a first order type of improvement over $\hat{f}_n(\cdot;m)$.

<u>Theorem 2</u> Suppose the density f has Fourier coefficients

$$\phi_j = e^{-aj} L(j) \ , \ j = 1, 2, \ldots, \text{ where}$$

L is a slowly varying function such that, for some $\delta>0$, the quantity $b_m(j) = \exp(\delta m)[L(j+m+1)/L(m)-1]$ satisfies

$$\lim_{m\to\infty} \sum_{j=0}^{\infty} b_m^2(j) e^{-2aj} = b \ , \text{ with } \ 0 < b < \infty \ .$$

It then follows that

$$\lim_{n\to\infty} \frac{J(\hat{f}_n(\cdot;m_n^*,e^{-a}),f)}{J(\hat{f}_n(\cdot;m_n),f)} = \lim_{n\to\infty} m_n^*/m_n = a/(a+\delta),$$

where m_n^* and m_n are as in Theorem 1.

Proof: From (3.1),

$$B(\hat{f}_n(\cdot;m,e^{-a}),f) = (2/\pi)\ \phi_m^2 \sum_{j=m+1}^{\infty} e^{-2a(j-m)}(L(j)/L(m)-1)^2$$

$$= (2/\pi)\ \phi_m^2\ e^{-2a}\ e^{-2\delta m} \sum_{j=0}^{\infty} b_m^2(j) e^{-2aj}$$

$$= (2/\pi)\, e^{-2(a+\delta)m}\, L^2(m) b e^{-2a}$$

$$+\ o(\phi_m^2 e^{-2\delta m}) \ .$$

Using Lemma 1 it now follows that

$$J(\hat{f}_n(\cdot;m,e^{-a}),f) \sim (1/\pi)\, [m/n + 2e^{-2(a+\delta)m}\, L^2(m) b e^{-2a}] \ .$$

Proceeding as in the proof of Theorem 1,

$$J(\hat{f}_n(\cdot;m_n^*,e^{-a}),f) \sim (1/\pi)(m_n^*/n) = \log n(2\pi(\delta+a)n)^{-1} + o(\log n/n).$$

The result follows after similarly analyzing the MISE of $\hat{f}_n(\cdot;m)$.

In Hart and Gray (1985), it was claimed that ARMA estimators are often more parsimonious than Fourier series estimators. Evidence of this is seen in Theorem 2. In fact, under the conditions of this theorem, the limits of $J(\hat{f}_n(\cdot;m_n^*,e^{-a}),f)/J(\hat{f}_n(\cdot;m_n),f)$ and m_n^*/m_n are one and the same, with both being less than 1. An example of ϕ_j's satisfying the conditions of Theorem 2 is

$$\phi_j = 1/\cosh(aj) \ , \ j = 0,1,\ldots \ .$$

It is easily verified that the value of δ for these ϕ_j's is $2a$, and so the MISE of $\hat{f}_n(\cdot;m_n^*,e^{-a})$ is asymptotically only 1/3 that of $\hat{f}_n(\cdot;m)$. Furthermore, this savings is obtained even though the ARMA estimator uses, in the limit, only 1/3 as many Fourier coefficients as does $\hat{f}_n(\cdot;m)$.

Also of interest is comparing ARMA estimators to general Fourier series estimators of the form

$$\hat{f}_w(x) = (1/\pi)\ (1+2\sum_{j=1}^{\infty} w_n(j)\hat{\phi}_j \cos jx) \ . \qquad (4.4)$$

Watson (1969) has shown that the minimum MISE among such estimators is

$$J_{n,opt} = (2/\pi) \sum_{j=1}^{\infty} \frac{\phi_j^2 \operatorname{var}(\cos jX_1)}{\operatorname{var}(\cos jX_1) + n\phi_j^2} \quad . \tag{4.5}$$

The following theorem shows that under the right conditions an ARMA estimator is better than any estimator of the form (4.4).

<u>Theorem 3</u> Suppose that in addition to the conditions of Theorem 2

$$|L(j)| < A < \infty \quad \text{for all } j.$$

It then follows that

$$\lim_{n\to\infty} J(\hat{f}_n(\cdot;m_n^*,e^{-a}),f)/J_{n,opt} = a/(a+\delta),$$

where m_n^* and δ are as in Theorem 2.

Proof: It is sufficient to show that $J_{n,opt} \sim \log n/(2na\pi)$. To do this, we proceed as in the proof of a theorem in Section 4 of Watson and Leadbetter (1963). If the ϕ_j's are absolutely summable, it is easy to show that

$$J_{n,opt} = (2/\pi) \sum_{j=1}^{\infty} \frac{\phi_j^2}{1+2(n-1)\phi_j^2} + O(1/n) \quad .$$

We have

$$\left| \sum_{j=1}^{\infty} \frac{\phi_j^2}{1+2(n-1)\phi_j^2} - \sum_{j=1}^{\infty} \frac{e^{-2aj}}{1+2(n-1)e^{-2aj}} \right| \le$$

$$(1+A^2) \sum_{j=1}^{\infty} \frac{e^{-2aj}}{[1+2(n-1)\phi_j^2][1+2(n-1)e^{-2aj}]} \le$$

$$(1+A^2)\{(n-1)^{-1}\sum_{j=1}^{c_n}\frac{1}{[1+2(n-1)\phi_j^2]}+\sum_{j=c_n+1}^{\infty}e^{-2aj}\}\ ,$$

where c_n is the greatest integer $\leq \log(n-1)/(2a)$.

Now, $\sum_{j=c_n+1}^{\infty} e^{-2aj} \leq (1-e^{-2a})^{-1}(n-1)^{-1}$. Also, since L^2 is slowly varying, we have $jL^2(j)\to\infty$ as $j\to\infty$ (see Seneta, p. 18). Hence, there exists j_0 such that for all $j \geq j_0$

$$[1+2(n-1)\phi_j^2]^{-1} < [1+2(n-1)j^{-1}e^{-2aj}C]^{-1} \quad (C > 0)$$

and $t^{-1}e^{-2at}$ is monotone decreasing on (j_0,∞). It follows that for all n sufficiently large

$$(n-1)^{-1}\sum_{j=1}^{c_n}[1+2(n-1)\phi_j^2]^{-1} < (n-1)^{-1}\sum_{j=j_0}^{c_n}[1+2(n-1)j^{-1}e^{-2aj}C]^{-1}$$

$$+O(n^{-2})$$

$$\leq (n-1)^{-1}\int_{j_0}^{\log(n-1)/2a}[1+2(n-1)t^{-1}e^{-2at}C]^{-1}dt + O(n^{-2})$$

$$\leq \frac{\log(n-1)}{2a(n-1)}\int_0^1 [1+4aCu^{-1}e^{-u\log(n-1)}\frac{(n-1)}{\log(n-1)}]^{-1}du + O(n^{-2})\ .$$

By dominated convergence, the previous integral tends to 0 as $n\to\infty$. The proof of the theorem is completed by showing that

$$\frac{n}{\log n}\sum_{j=1}^{\infty}\frac{e^{-2aj}}{1+2(n-1)e^{-2aj}} \to (4a)^{-1} \text{ as } n\to\infty.$$

The previous theorem shows that ARMA estimators are fundamentally different than Fourier series estimators of the form (4.4). The full extent to which this is true will become apparent in the next two sections.

Before proceeding, we also note that a result analogous to Theorem 1 can undoubtedly be obtained for densities satisfying $\phi_j \sim c\exp(-aj^{\gamma})$, $0 < \gamma < 1$. For these densities, Hart and Gray (1985) show that if $\alpha_m = \phi_{m+1}/\phi_m$, then $B(\hat{f}_n(\cdot;m,\alpha_m),f)/B(\hat{f}_n(\cdot;m),f)\to 0$ as $m\to\infty$.

5. Densities With Regularly Varying Fourier Coefficients

In a certain sense, the results in Section 4 are not particularly surprising. The Fourier coefficients of $\hat{f}_n(\cdot;m,e^{-a})$ are reasonably well matched to those of the underlying density, and so one might expect the ARMA estimator to have much smaller bias than $\hat{f}_n(\cdot;m)$. Of more interest is to investigate the robustness of ARMA estimators to departures from the approximate exponential model for the ϕ_j's.

In this section we study the case

$$\phi_j = R_{-\rho}(j) \text{ , } j = 1,2,\ldots, \tag{5.1}$$

where $R_{-\rho}$ is a continuous, regularly varying function of index $-\rho$ and $\rho>1/2$. Included in such cases are densities with Fourier coefficients

$$\phi_j = (1+(j/a)^{\rho})^{-1} \text{ , } j=0,1,\ldots, \tag{5.2}$$

where $1/2 < \rho \leq 2$. More generally, any set of square summable ϕ_j's that decay algebraically (see Watson & Leadbetter 1963 and Davis 1977) satisfy (5.1).

We here present a lemma concerning the integrated squared bias of ARMA estimators when (5.1) holds.

<u>Lemma 3</u> Let the ϕ_j's be as in (5.1), and suppose that $m(1-\alpha_m) \to c > 0$ as $m \to \infty$. Then

$$\sum_{j=m+1}^{\infty} (\phi_j - \phi_m \alpha_m^{j-m})^2 = m\phi_m^2 \int_0^\infty ((1+y)^{-\rho} - e^{-cy})^2 dy + o(m\phi_m^2).$$

Proof: Since the ϕ_j's are regularly varying, $\sum_{j=m+1}^{\infty} \phi_j^2/(m\phi_m^2) \to \int_0^\infty (1+y)^{-2\rho} dy$ as $m \to \infty$. Also, $\phi_m^2 \sum_{j=m+1}^{\infty} \alpha_m^{2(j-m)}/(m\phi_m^2) = \alpha_m^2\, m^{-1}(1-\alpha_m^2)^{-1} \to 1/(2c) = \int_0^\infty e^{-2cy} dy$. There exist (see Seneta, pp. 19-20) functions $\underline{\phi}$ and $\bar{\phi}$, defined on $(0,\infty)$, with the properties $\underline{\phi}(t) \sim \bar{\phi}(t)$ as $t \to \infty$ and $\underline{\phi}(t) \le \phi_{[t]+1} \le \bar{\phi}(t)$ for all $t>0$. By assumption,

$$1-(1+\varepsilon)c/m \le \alpha_m \le 1 - (1-\varepsilon)c/m$$

for $0 < \varepsilon < 1$ and all m sufficiently large. It follows that

$$\lim_{m\to\infty} \sup \sum_{j=m+1}^{\infty} \phi_j \, \alpha_m^{j-m}/(m\phi_m) \le$$

$$\lim_{m\to\infty} \int_m^\infty (m\phi_m)^{-1} \bar{\phi}(t)(1-(1-\varepsilon)c/m)^{t-m} dt =$$

$$\lim_{m\to\infty} \int_1^\infty (\bar{\phi}(mu)/\phi_m)(1-(1-\varepsilon)c/m)^{m(u-1)} du.$$

Making use of dominated convergence, the last limit is

$$\int_1^\infty u^{-\rho} e^{-(1-\varepsilon)c(u-1)} du .$$

Obtaining a similar lower bound, and using the fact that ε may be taken arbitrarily small, we have

$$\lim_{m\to\infty} \sum_{j=m+1}^{\infty} \phi_j \alpha_m^{j-m} /(m\phi_m) = \int_0^\infty (1+y)^{-\rho} e^{-cy} dy.$$

The lemma now follows upon combining the previous results.

Some remarks are in order here. First, we see that the choice of α_m in Lemma 3 is such that $\alpha_m \sim 1$. This is required since the ϕ_j's in (5.1) decay to zero more slowly than do those in the previous section. It is straightforward to show that if (5.1) holds and $|\alpha_m|$ is bounded away from 1, the limiting integrated squared bias of $\hat{f}_n(\cdot;m,\alpha_m)$ is the same as that of $\hat{f}_n(\cdot;m)$.

Now, since the ϕ_j's are regularly varying with index $-\rho < -1/2$, it follows that

$$\sum_{j=m+1}^{\infty} \phi_j^2/(m\phi_m^2) \to (2\rho-1)^{-1} = \int_0^{\infty}(1+y)^{-2\rho}dy. \tag{5.3}$$

Hence, the integrated squared biases of $\hat{f}_n(\cdot;m)$ and $\hat{f}_n(\cdot;m,\alpha_m)$ (with α_m as in Lemma 3) are both of order $m\phi_m^2$ as $m\to\infty$. However, by choosing α_m such that $m(1-\alpha_m)\to\rho$,

$$B(\hat{f}_n(\cdot;m,\alpha_m),f)/B(\hat{f}_n(\cdot;m),f) \to (2\rho-1) \int_0^{\infty} ((1+y)^{-\rho}-e^{-\rho y})^2dy < 1.$$

Note, though, that with $m(1-\alpha_m)\to\rho$, the contribution of α_m to the integrated variance of the ARMA estimator is not insignificant. The smaller bias of the ARMA estimator does not, then, immediately imply that its MISE is smaller than that of $\hat{f}_n(\cdot;m)$. This point is investigated in the next theorem.

<u>Theorem 4</u> Let the Fourier coefficients of the density f be as in (5.1), and suppose that $m(1-\alpha_m)\to c > 0$ as $m\to\infty$. Then, defining

$$I_{\rho,c} = \int_0^{\infty}((1+y)^{-\rho}-e^{-cy})^2dy,$$

$$J(\hat{f}_n(\cdot;m,\alpha_m),f) = \pi^{-1}[(m/n)(1+(2c)^{-1}) + 2m\phi_m^2\, I_{\rho,c}]$$

$$+ o(m/n + m\phi_m^2), \text{ and}$$

$$J(\hat{f}_n(\cdot;m),f) = \pi^{-1}[m/n + 2m\phi_m^2(2\rho-1)^{-1}]$$

$$+ O(\sqrt{m}/n) + o(m\phi_m^2) .$$

Furthermore, if m_n^* and m_n are the minimizers of, respectively, $[(m/n)(1+(2c)^{-1}) + 2(2\rho-1)I_{\rho,c} \sum_{j=m+1}^{\infty} \phi_j^2]$ and $J(\hat{f}_n(\cdot;m),f)$, then

$$J(\hat{f}_n(\cdot;m_n^*,\alpha_{m_n^*}),f)/J(\hat{f}_n(\cdot;m_n),f) \sim$$

$$(m_n^*/m_n)(1+(2c)^{-1}) + [(2\rho-1)I_{\rho,c}]^{1/(2\rho)}(1+(2c)^{-1})^{1-1/(2\rho)}$$

as $n\to\infty$.

Proof: The two MISE expressions follow from (3.1) upon applying the condition $m(1-\alpha_m)\to c$, Lemmas 1 and 3, and expression (5.3). Now, using (5.3) and the first part of Theorem 4,

$$J(\hat{f}_n(\cdot;m,\alpha_m),f) \sim \pi^{-1}[(m/n)(1+(2c)^{-1}) + (2\rho-1)\ I_{\rho,c}\sum_{j=m+1}^{\infty}\phi_j^2].$$

If m_n^* is the minimizer of this last expression, then the continuity of $R_{-\rho}$ implies that $R_{-\rho}^2(t_n) = A^{-1}n^{-1}$ for some $t_n \in [m_n^*, m_n^*+1]$, where $A = 2(2\rho-1)I_{\rho,c}(1+(2c)^{-1})^{-1}$. By 5°, p. 21 of Seneta, there exists a regularly varying function r of index $1/(2\rho)$ such that

$$r(1/R_{-\rho}^2(t)) \sim t \quad \text{as} \quad t\to\infty.$$

Since $t_n \to \infty$, t_n, and hence m_n^*, are asymptotic to $r(An)$. We now have

$$J(\hat{f}_n(\cdot;m_n^*,\alpha_{m_n^*}),f) \sim \pi^{-1}(1+(2c)^{-1})2\rho(2\rho-1)^{-1}r(An)/n.$$

In a similar way, it can be shown that

$$J(\hat{f}_n(\cdot;m_n),f) \sim \pi^{-1}2\rho(2\rho-1)^{-1}r(2n)/n.$$

The rest of Theorem 4 now follows upon using the facts $m_n^* \sim r(An)$, $m_n \sim r(2n)$, and r is regularly varying of index $1/(2\rho)$.

To determine the amount of improvement (if any) that is possible with an ARMA estimator, it would be desirable to choose c to minimize the limiting ratio of MISEs in Theorem 4. This minimization problem appears to be intractable analytically, although one could certainly determine the best c numerically for any given ρ. In the following theorem it is shown that the choice $c = \rho$ is such that the limiting ratio in Theorem 4 is less than 1 for each $\rho > 1/2$. Hence, under the conditions of Theorem 4, it is always possible to obtain asymptotically smaller MISE with an ARMA estimator than with $\hat{f}_n(\cdot;m)$.

<u>Theorem 5</u> Let the conditions of Theorem 4 hold with $c = \rho$. Then, for each $\rho > 1/2$,

$$\lim_{n\to\infty} J(\hat{f}_n(\cdot;m_n^*,\alpha_{m_n^*}),f)/J(\hat{f}_n(\cdot;m_n),f) =$$

$$B_\rho = [(2\rho-1)I_{\rho,\rho}]^{1/(2\rho)}(1+(2\rho)^{-1})^{1-1/(2\rho)} < 1.$$

Proof: Using the fact that $(1+y)^{-\rho}\, e^{-\rho y} > e^{-2\rho y}$ for $y>0$, it is easy to show that $(2\rho-1)I_{\rho,\rho} < 1/(2\rho)$. Hence, $B_\rho < (2\rho)^{-1/(2\rho)}(1+(2\rho)^{-1})^{1-1/(2\rho)}$, which is less than or equal to 1 iff $2\rho\log(2\rho) - (2\rho-1)\log(2\rho+1) \geq 0$. Defining

$$g(x) = x\log x - (x-1)\log(x+1) \text{ for } x \geq 1,$$

it is now sufficient (since $g(1) = 0$) to show that $g'(x) > 0$ for $x > 1$. We have $g'(x) = 2/(x+1)-\log(1+x^{-1})$ with $g'(1) = 1 - \log 2 > 0$. Since $g'(x) \to 0$ as $x \to \infty$, the result will follow if g' is monotone decreasing on $(1,\infty)$. We have $g''(x) = -2(x+1)^{-2} + x^{-1} - (x+1)^{-1}$, which is less than 0 on $(1,\infty)$, and the proof is complete.

Recalling the e_1-transform motivation for ARMA estimators, it is of interest to determine if $\alpha_m = \phi_{m+1}/\phi_m$ satisfies $m(1-\alpha_m)\to\rho$ when (5.1) holds. As shown by Seneta, pp. 2-7, $R_{-\rho}$ is regularly varying with index $-\rho$ if and only if it satisfies, for all t sufficiently large,

$$R_{-\rho}(t) = t^{-\rho}\exp(\eta(t) + \int_B^t \varepsilon(x)x^{-1}dx), \tag{5.4}$$

where η is a bounded, measurable function on $[B,\infty)$ such that $\eta(t)\to c_0$ as $t\to\infty$ ($|c_0|<\infty$), and ε is a continuous function on $[B,\infty)$ such that $\varepsilon(t)\to 0$ as $t\to\infty$. Now, if the function η in (5.4) is replaced by a constant, it is easy to show that $m(1-R_{-\rho}(m+1)/R_{-\rho}(m))\to\rho$ as $m\to\infty$. Therefore, Theorem 5 implies that under a slightly stronger condition on the ϕ_j's than imposed by (5.1), the MISE of an ARMA estimator with $\alpha_m = \phi_{m+1}/\phi_m$ is asymptotically smaller than that of $\hat{f}_n(\cdot;m)$. Since (5.4) does not decay geometrically, the last result gives a good indication of the versatility of the e_1-transform based ARMA estimator.

We close this section with a corollary that extends the results of Sections 4 and 5.

Corollary 1 Suppose the ϕ_j's in Theorems 1-5 are replaced by $(-1)^j\phi_j$ for $j = 1,2,\ldots$, and the respective values for α are replaced by $-\alpha$. Then the results of the five theorems are unchanged.

Proof: Considering expression (3.1), it is clear that $J(\hat{f}_n(\cdot;m,\alpha),f) = J(\hat{f}_n(\cdot;m,-\alpha),f^*)$, where f and f^* have Fourier coefficients, respectively, ϕ_j and $(-1)^j\,\phi_j$, $j = 1,2,\ldots$.

6. Densities That Are Smooth at One Endpoint but Not the Other

In this section it is shown that the results of Section 5 are applicable under simple qualitative conditions concerning the smoothness of f. When these conditions hold, it is found that a result like that in Theorem 5 remains valid even when $\hat{f}_n(\cdot;m)$ is replaced by the very best estimator of the form (4.4). To obtain this somewhat surprising result, we first state the following lemma. The proof is analogous to that of the first theorem in Section 3 of Watson and Leadbetter (1963), and is thus omitted.

Lemma 4 Suppose that for some $\rho > 1/2$ $j^\rho|\phi_j| \to K^{1/2} > 0$ as $j \to \infty$. Then

$$J_{n,opt} \sim n^{1/(2\rho)-1}\pi^{-1}(2K)^{1/(2\rho)}\int_0^\infty (1+t^{2\rho})^{-1}dt,$$

where $J_{n,opt}$ is defined by (4.5).

The following theorem will allow us to obtain the most interesting results of this paper.

Theorem 6 Suppose that f has $2k$ derivatives $(k \geq 1)$ on $[0,\pi]$ with $f^{(2k)}$ square integrable. (The quantities $f^{(r)}(0)$ and $f^{(r)}(\pi)$ are defined to be $f^{(r)}(0+)$ and $f^{(r)}(\pi-)$, respectively.) If $f^{(r)}(0) = f^{(r)}(\pi) = 0$ for

$r = 1, 3, \ldots, 2k-3$, $f^{(2k-1)}(\pi) = 0$ and $f^{(2k-1)}(0) \neq 0$, then (with $m(1-\alpha_m) \to 2k$)

$$J(\hat{f}_n(\cdot;m_n^*,\alpha_{m_n^*}),f) \sim n^{1/(4k)-1}\pi^{-1}\left|f^{(2k-1)}(0)\right|^{(1/2k)}C_k, \tag{6.1}$$

$$\lim_{n\to\infty} J(\hat{f}_n(\cdot;m_n^*,\alpha_{m_n^*}),f)/J(\hat{f}_n(\cdot;m_n),f) =$$

$$[(4k-1)I_{2k,2k}]^{1/(4k)}(1+(4k)^{-1})^{1-1/(4k)} < 1, \text{ and} \tag{6.2}$$

$$\lim_{n\to\infty} J(\hat{f}_n(\cdot;m_n^*,\alpha_{m_n^*}),f)/J_{n,opt} =$$

$$C_k[2^{1/(4k)}\pi/(4k\sin(\pi/(4k)))]^{-1} < 1 , \tag{6.3}$$

where

$$C_k = [(4k+1)/(4k-1)]^{1-1/(4k)}(8kI_{2k,2k})^{1/(4k)} .$$

If $f^{(r)}(0) = f^{(r)}(\pi) = 0$ for $r = 1, 3, \ldots, 2k-3$, $f^{(2k-1)}(0) = 0$, $f^{(2k-1)}(\pi) \neq 0$, and $m(1+\alpha_m) \to 2k$, then the result of the theorem is the same except that $f^{(2k-1)}(0)$ is replaced by $f^{(2k-1)}(\pi)$.

Proof: Using integration by parts we have

$$(-1)^k\phi_j = j^{-2k}[f^{(2k-1)}(0) + \phi_{j,2k}] \text{ , where}$$

$$\phi_{j,2k} = \int_0^\pi f^{(2k)}(u)\cos ju\,du.$$

Since $f^{(2k)}$ is square integrable, $\phi_{j,2k} \to 0$ as $j \to \infty$. It follows that the ϕ_j's are regularly varying with index $-2k$ and that $j^{2k}|\phi_j| \to |f^{(2k-1)}(0)|$ as

$j\to\infty$. Results (6.1) and (6.2) now follow upon applying Theorems 4 and 5, respectively. The limit in (6.3) is a consequence of Theorem 4, Lemma 4, and the fact that

$$\int_0^\infty (1+t^{4k})^{-1}dt = \pi[4k\sin(\pi/(4k))]^{-1} .$$

It is argued in the Appendix that the limit in (6.3) is less than 1. The case where $f^{(2k-1)}(0) = 0$ and $f^{(2k-1)}(\pi) \neq 0$ follows from Corollary 1.

The special case $k = 1$ in Theorem 6 is worth discussing. In this case, we have either $f'(0) \neq 0$, $f'(\pi) = 0$ or $f'(0) = 0$, $f'(\pi) \neq 0$. In either situation, the density tends smoothly to its limit at one endpoint but not at the other; exponential-like densities are one example of such behavior. The poor performance of the Fourier series estimator $\hat{f}_n(0;m)$ when $f'(0) \neq 0$ has been noted by Buckland (1985) in the setting of line transect sampling. Theorem 6 shows that in such cases the MISE of an ARMA estimator is (for large enough n) smaller than that of the very best Fourier series estimator of the form (4.4) (which, of course, includes $\hat{f}_n(\cdot;m)$ as a special case).

Table 1 shows how much of an improvement in MISE is obtained with the ARMA estimator in certain cases where the ϕ_j's decay algebraically. The cases $\rho = 2, 4, 10, 20$ correspond to $k = 1, 2, 5, 10$ in Theorem 6. Note that when $\rho = 2$ (the situation discussed in the previous paragraph) the asymptotic MISE of the ARMA estimator is only about 64% and 77% of that for, respectively, $\hat{f}_n(\cdot;m)$ and the optimum Fourier series estimator. It is also interesting that the ARMA estimator uses, in the limit, only 51% as many Fourier coefficients as does $\hat{f}_n(\cdot;m)$.

7. Choice of Smoothing Parameter by Cross-Validation

In the practice of density estimation, one must usually select smoothing parameters via some data-driven method. As a result, data-based density estimators are not as efficient (at least in small samples) as theory would suggest. Hence, it is not realistic to expect that the MISE improvement discussed in previous sections is fully attainable in practice, except perhaps in very large samples. However, some improvement in small samples seems likely if the ARMA estimator's smoothing parameter is reasonably chosen. This is borne out in the simulation study of this section.

One means of choosing the smoothing parameter, (m,α), of the ARMA estimator is the cross-validatory method introduced by Rudemo (1982) and Bowman (1984). In our setting, this method chooses the pair $(\hat{m},\hat{\alpha})$ that minimizes

$$\hat{R}(m,\alpha) = \int_0^\pi \hat{f}_n^2(x;m,\alpha)dx - (2/n) \sum_{i=1}^n \hat{f}_{n,i}(X_i;m,\alpha), \tag{7.1}$$

where $\hat{f}_{n,i}$ indicates the estimator calculated by deleting the data value X_i. Rudemo (1982) showed that $\hat{R}(m,\alpha)$ is an unbiased estimator of the risk

$$R(m,\alpha) = J(\hat{f}_n(\cdot;m,\alpha),f) - \int_0^\pi f^2(x)dx.$$

A number of results now exist showing that density estimates chosen by cross-validation are asymptotically efficient; see, for example, Hall (1983, 1985), Stone (1984), and Hall and Marron (1985).

To investigate the behavior of cross-validated ARMA estimates, a small simulation study was conducted. The density considered was

$$f(x) = 2e^{-2x}(1+e^{-4(\pi-x)})(1-e^{-4\pi})^{-1}, \quad 0 \le x \le \pi, \tag{7.2}$$

which has algebraic Fourier coefficients

$$\phi_j = (1+(j/2)^2)^{-1}, \quad j = 1,2,\ldots \quad .$$

This density is of interest since it has $f'(0) = -2(1-e^{-4\pi})$ and $f'(\pi) = 0$, and thus satisfies the conditions of Theorem 6.

For (7.2) and $n = 50$, the minimum MISE among ARMA estimators is .00633. This minimum occurs at $(m,\alpha) = (1, .64)$. Among Fourier series estimators $\hat{f}_n(\cdot;m)$, the optimum m and MISE are 5 and .04160. Since .00633/.0416 = .152, we see that the asymptotic relative efficiency of .64 from Table 1 understates the improved efficiency of the optimum ARMA estimator at $n = 50$.

Twenty independent random samples of size $n = 50$ were generated from the density (7.2). This was done by generating values from the exponential density $g(x) = 2e^{-2x} I_{(0,\infty)}(x)$, and using the fact that, if Y has density g, then

$$X = Y I_{(0,\pi)}(Y) + \sum_{j=1}^{\infty} (Y-2j\pi)[I_{(2j\pi,(2j+1)\pi)}(Y) - I_{((2j-1)\pi,2j\pi)}(Y)]$$

has density (7.2). Since little of the mass of g is larger than π, the graphs of f and g on $(0,\pi)$ are virtually indistinguishable.

For each of the twenty data sets, the minimizer of $\hat{R}$ (defined by (7.1)) for $0 \leq \alpha < 1$, $1 \leq m \leq 20$ was approximated. Also, the minimizer of $\hat{R}(m,0)$ for $1 \leq m \leq 20$ was determined. This latter value of m is simply a cross-validatory choice of the smoothing parameter of $\hat{f}_n(\cdot;m)$ (see Hart 1985 and Diggle and Hall 1986 for more on this subject). The integrated squared errors of the cross-validated ARMA and Fourier series estimates were determined for each data set. Denote these two ISEs I_A and I_F, respectively.

The results of the simulation are summarized in Table 2. The fact that average I_A was a bit larger than average I_F is misleading. In 16 of the 20 cases, the ratio I_A/I_F was between .151 and .690. Note the trimmed mean and the confidence interval for the median of I_A/I_F in Table 2. These more accurately reflect the overall performance of the two cross-validated density estimates. From the 16 cases in which I_A was less than I_F, a typical comparison of the two estimates is given in Figure 1. The qualitative improvements obtained with the ARMA estimate are a better estimate of $f(0)$ and an absence of spurious bumps. See Hart and Gray (1985) for further discussion of the qualitative properties of ARMA approximators.

It is also important to point out what happened in the four cases where $I_A > I_F$. Figure 2 is a plot of the 20 values of (m,α) chosen by cross-validation. In the cases where $I_A > I_F$, cross-validation chose too large an α and/or too large an m. When α is too near 1, the ARMA estimate tends to be too large near 0, thus inflating the estimate's integrated squared error. That cross-validation would occasionally choose too rough an ARMA estimate is not unexpected. It is well known from other settings that in small samples cross-validation tends to drastically undersmooth around 5-20% of the time (see, e.g., Hart 1985).

Clearly, further experimentation is needed to determine how efficiently cross-validation smooths ARMA estimates. Some very recent work of Scott and Terrell (1986) shows that a biased version of cross-validation provides more efficient kernel estimators in moderate and large samples. Such an idea is also worth pursuing in the setting of ARMA estimators. For example, a modification of $\hat{R}(m,\alpha)$ that places a more severe penalty on large values of $|\alpha|$ would discourage the occasional undersmoothing observed in the simulation study.

8. Concluding Remarks

We have shown that under a variety of conditions, a density estimator having the rational form of an ARMA spectrum can yield an improvement in MISE over the very best Fourier series estimator of the form (4.4). Furthermore, if its smoothing parameter is chosen properly, the ARMA estimator never has larger MISE than that of the simple Fourier series estimator $\hat{f}_n(\cdot;m)$. Although further work is needed on the data-based smoothing of ARMA estimates, the results of Section 7 are fairly encouraging with regard to cross-validated smoothing.

The ideas of this paper could also be applied to other density estimators of the type

$$\hat{f}_m(x) = (1/\pi)\ [1+2 \sum_{j=1}^{m} w_j\ \hat{\phi}_j \cos jx]\ .$$

The form (2.4) immediately suggests "jackknifed" versions of $\hat{f}_m$. So long as the truncation bias of $\hat{f}_m$ is not dominated by the bias due to the w_j's, results analogous to, for example, Theorem 5 could be established for a jackknifed $\hat{f}_m$.

Finally, our results could also be extended somewhat by considering estimators based on a sine-cosine basis and allowing α to be complex-valued. In fact, the results of Hart and Gray (1985) apply to the bias of such estimators.

Acknowledgements

The author expresses his deep appreciation to H. L. Gray, whose work in numerical analysis was the stimulus for this paper. Also, the author benefitted greatly from the insight of Daren B. H. Cline, who provided a proof of Lemma 3.

Appendix

Proof of (6.3): As stated in the proof of Theorem 6, we need only show that the limit in (6.3) is less than 1. For $k = 1$ or 2, the inequality follows from Table 1. For $k = 3,4,\ldots$, it is sufficient to show that $C_k 2^{-1/(4k)} \leq 1$, since $0 < x^{-1}\sin x < 1$ for $0 < x \leq \pi/12$. Now,

$$C_k 2^{-1/(4k)} = 4k(4k-1)^{-1}(1+(4k)^{-1})^{1-1/(4k)}[(4k-1)I_{2k,2k}]^{1/(4k)}$$

$$<4k(4k-1)^{-1}(1+(4k)^{-1})^{1-1/(4k)}(4k)^{-1/(4k)}\ , \text{ as}$$

in the proof of Theorem 5. The last quantity is less than or equal to 1 for $k = 3, 4,\ldots$ if

$$g(x) = x\log(x-1) - (x-1)\log(x+1) \geq 0$$

for $x \geq 12$. The rest of the proof proceeds exactly as in the proof of Theorem 5. (Note that $g(12) > 0$.)

Calculation of numbers in Table 1: The requisite quantities can all be calculated analytically except for

$$(2\rho-1)I_{\rho,\rho} = 2 - (2\rho)^{-1} - 2(2\rho-1)e^{\rho}E_{\rho}(\rho)\ ,$$

where $E_k(y) = \int_1^\infty e^{-yt}t^{-k}dt$ is the so called exponential integral. The table values were obtained by using the approximation of either $E_\rho(\rho)$ or $e^{\rho}E_{\rho}(\rho)$ given in Abramowitz and Stegun (1972).

References

Abramowitz, M. and Stegun, I. A. (1972). Handbook of Mathematical Functions. Dover, New York.

Aitken, A.C. (1926). On Bernoulli's numerical solution of algebraic equations. Proc. Roy. Soc. Edinburgh A 46 289-305.

Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. Biometrika 65 521-528.

Buckland, S.T. (1985). Perpendicular distance models for line transect sampling. Biometrics 41 177-195.

Carmichael, J.P. (1984). Consistency of an autoregressive density estimator. Mathematische Operationsforschung und Statistik, Series Statistics 15 383-387.

Cencov, N.N (1962). Evaluation of an unknown distribution density from observations. Soviet Mathematics 3 1559-1562.

Crain, B.R., Burnham, K.P., Anderson, D.R., and Laake, J.L. (1979). A Fourier series estimator of population density for line transect sampling. Biometrical Journal 21 731-748.

Davis, K.B. (1977). Mean integrated square error properties of density estimates. Ann. Statist. 5 530-535.

Diggle, P.J. and Hall, P. (1986). The selection of terms in an orthogonal series density estimator. J. Amer. Statist. Assoc. 81 230-233.

Gates, C.E. and Smith, P.W. (1980). An implementaton of the Burnham-Anderson distribution-free method of estimating wildlife densities from line transect data. Biometrics 36 155-160.

Gray, H.L., Watkins, T.A., and Adams, J.E. (1972). On the jackknife statistic, its extensions, and its relation to e_n-transformations. Ann. Math. Statist. 43 1-30.

Gray, H.L. (1985). On a unification of bias reduction and numerical approximation. To appear in Graybill Festschrift, North Holland.

Hall, P. (1983a). Measuring the efficiency of trigonometric series estimates of a density. J. Multivariate Anal. 13 234-256.

Hall, P. (1983b). Large sample optimality of least squares cross-validation in density estimation. Ann. Statist. 11 1156-1174.

Hall, P. (1985). Smoothing orthogonal series density estimators, I. Manuscript.

Hall, P. and Marron, J.S. (1985). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. Center for Stochastic Processes Technical Report no. 94.

Hart, J.D. (1985). On the choice of a truncation point in Fourier series density estimation. J. Statist. Comput. Simulation 21 95-125.

Hart, J.D. and Gray, H.L. (1985). The ARMA method of approximating probability density functions. J. Statist. Planning and Inference 12 137-152.

Kronmal, R.A. and Tarter, M.E. (1968). The estimation of probability densities and cumulatives by Fourier series methods. J. Amer. Statist. Assoc. 63 925-952.

Morton, M.J. and Gray, H.L. (1984). The G-spectral estimator. J. Amer. Statist. Assoc. 79 692-701.

Parzen, E. (1979). Nonparametric statistical data modeling. J. Amer. Statist. Assoc. 74 105-121.

Rudemo, M. (1982). Empirical choice of histogram and kernel density estimators. Scand. J. Statist. 9 65-78.

Schucany, W.R., Gray, H.L. and Owen, D.B. (1971). On bias reduction in estimation. J. Amer. Statist. Assoc. 66 524-533.

Schuster, E.F. (1985). Incorporating support constraints into non-parametric estimators of densities. Commun. Statist. - Theor. Meth. 14 1123-1136.

Scott, D.W. and Terrell, G.R. (1986). Biased and unbiased cross-validation in density estimation. Stanford University Department of Statistics Technical Report No. 23.

Seneta, E. (1976). Regularly Varying Functions. Springer-Verlag, Berlin.

Shanks, D. (1955). Nonlinear transformation of divergent and slowly convergent sequences. J. Math. Phys. 34 1-42.

Stone, C.J. (1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist. 12 1285-1297.

Watson, G. S. and Leadbetter, M.R. (1963). On the estimation of the probability density, I. Ann. Math. Statist. 34 480-491.

Watson, G.S. (1969). Density estimation by orthogonal series. Ann. Math. Statist. 40 1496-1498.

Table 1. Asymptotic MISE of ARMA and Fourier Series Estimators for Densities With Algebraically Decaying Fourier Coefficients

(It is assumed that $j^{\rho}|\phi_j| \to K^{1/2}$ as $j \to \infty$.)

Type of Estimator

ρ	ARMA	Optimum Fourier Series	Fourier Series with 0-1 Weights
1	1.3579	1.5708	2
2	.8534	1.1107	4/3
3	.7968	1.0472	6/5
4	.7890	1.0262	8/7
10	.8335	1.0041	20/19
20	.8826	1.0010	40/39

Notes: For a given ρ and estimator, a table value is the limit of $n^{1-1/(2\rho)} B_\rho$ MISE, where $B_\rho = \pi(2K)^{-1/(2\rho)}$. Details of how the values were obtained are given in the Appendix. For a given ρ, the limiting ratio of ARMA truncation point (m_n^*) to Fourier series truncation point (m_n) is $(1+(2\rho)^{-1})^{-1} A_\rho/FS_\rho$, where A_ρ and FS_ρ are the table values in, respectively, the first and third columns above.

Table 2. Summary of Simulation Study

	Type of Estimate	
	ARMA	Fourier Series
Average ISE	.0542	.0522
Median ISE	.0183	.0354
Trimmed mean ISE	.0287	.0390

Notes: The trimmed means exclude the three (out of 20) largest values of ISE. A 95% confidence interval for the median of I_A/I_F is (.3607, .6581), where I_A and I_F are, respectively, the ISEs of cross-validated ARMA and Fourier series estimates.

Captions for Figures

Figure 1. Wrapped Exponential Density, ARMA Estimate, and Fourier Series Estimate. The solid curve is the density (7.2). The ARMA estimate has the higher value at 0 and no spurious bumps. The two estimates were calculated from the same set of data, each being fitted by cross-validation.

Figure 2. Distribution of Smoothing Parameters in Simulation Study. The smoothing parameters were chosen by cross-validation. The four largest values of m correspond to the only cases where the ISE of the ARMA estimate was larger than that of the Fourier series estimate. The MISE optimum value of (m,α) is $(1,.64)$.

Figure 1.

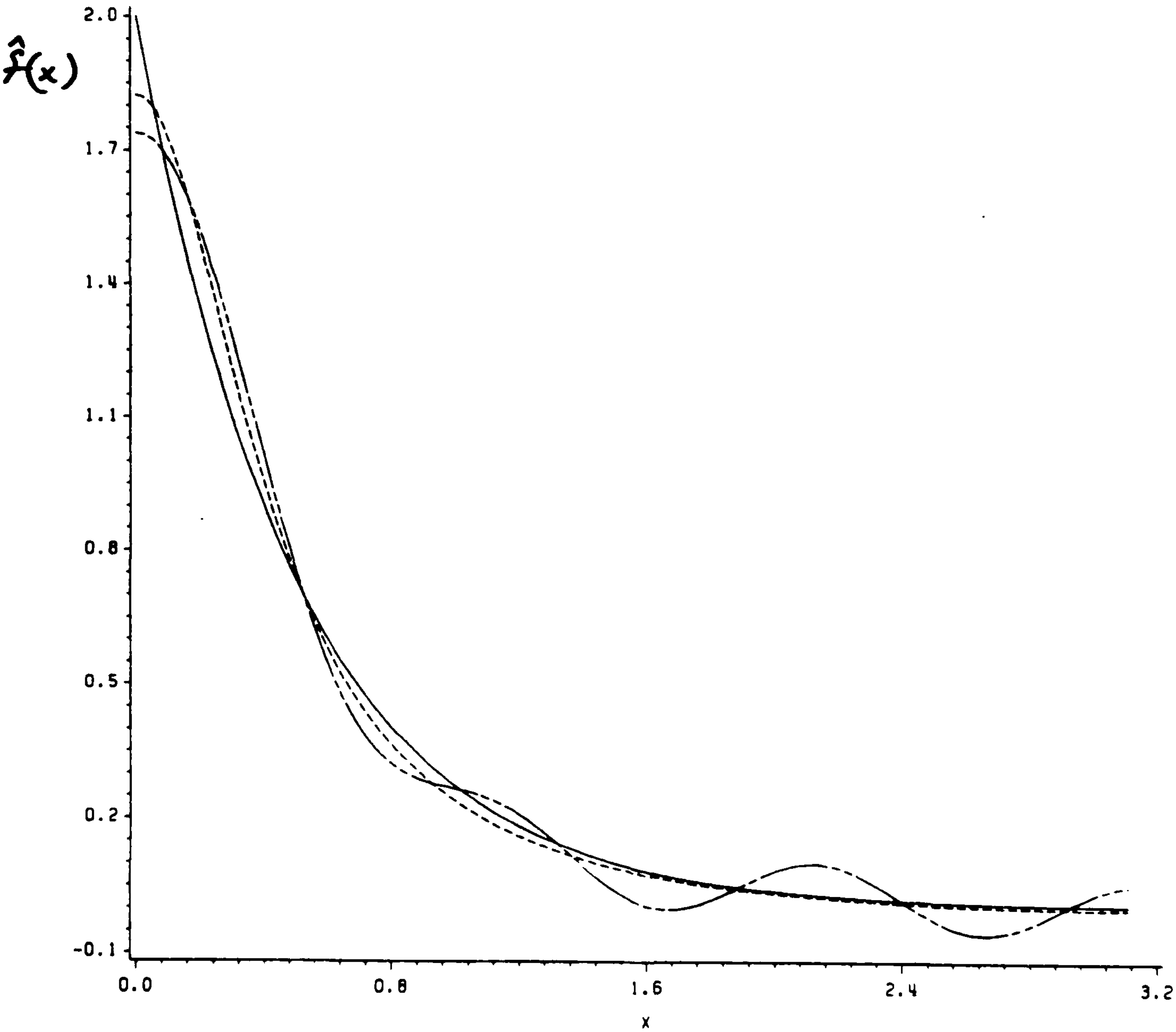
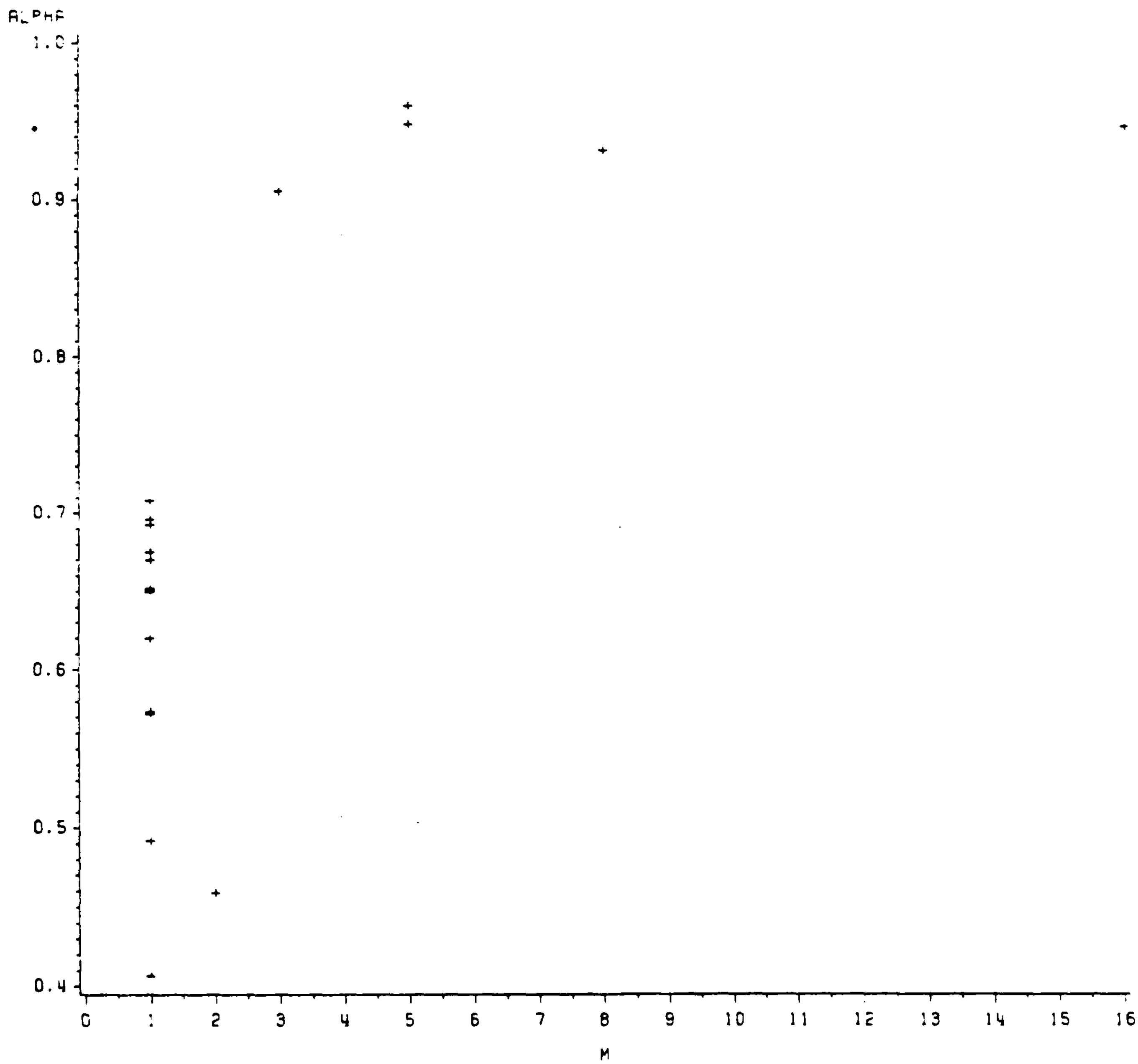


Figure 2.



END
8-87
DTIC